First-Digit Law in Nonextensive Statistics*

Lijing Shao[1] and Bo-Qiang Ma[1,2,†]

[1]*School of Physics and State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing 100871, China*

[2]*Center for High Energy Physics, Peking University, Beijing 100871, China*

Abstract

Nonextensive statistics, characterized by a nonextensive parameter q, is a promising and practically useful generalization of the Boltzmann statistics to describe power-law behaviors from physical and social observations. We here explore the unevenness of the first digit distribution of nonextensive statistics analytically and numerically. We find that the first-digit distribution follows Benford's law and fluctuates slightly in a periodical manner with respect to the logarithm of the temperature. The fluctuation decreases when q increases, and the result converges to Benford's law exactly as q approaches 2. The relevant regularities between nonextensive statistics and Benford's law are also presented and discussed.

PACS numbers: 02.50.Cw, 05.20.-y

* Published in Phys. Rev. E82, 041110 (2010)

†Corresponding author. Electronic address: mabq@pku.edu.cn

I. INTRODUCTION

The foundation of statistical physics was established at the end of 19th century by Maxwell, Boltzmann, and Gibbs [1–4]. The celebrated Maxwell-Boltzmann-Gibbs statistics underpins revolutionary concepts such as ergodicity and stosszahlansatz, which means that all admissible microstates are equiprobable over a long period of time. Cooperative effects and nonlinear dynamics are important in leading to sufficient statistics. The Boltzmann distribution has a distinguished exponential behavior proportional to $e^{-\beta E}$, where E is the energy, and $\beta = 1/k_B T$ is the reciprocal of the temperature T with the Boltzmann constant k_B.

Recently, it was found that the Boltzmann statistics follows a peculiar digit law [5], named Benford's law, which is also called the first-digit law or the significant digit law [6, 7]. The law states that the occurrence of the nonzero leftmost digit, *i.e.*, $1, 2, ..., 9$, of numbers from many real world sources is not uniformly distributed as one might naively expect, but instead, the nature favors smaller ones according to a logarithmic formula,

$$P_{\mathrm{Ben}}(d) = \log_{10}\left(1 + \frac{1}{d}\right), \quad d = 1, 2, ..., 9, \tag{1}$$

where $P_{\mathrm{Ben}}(d)$ is the probability of a number having the first digit d. The first-digit distribution of the Boltzmann statistics fluctuates slightly around Benford's law in a periodical manner with respect to the logarithm of the temperature of the system [5]. Moreover, two quantum distributions, *i.e.*, the Bose-Einstein statistics and the Fermi-Dirac statistics, are also proven to comply with Benford's law in analogous manners [5].

On the other hand, the natural and social numbers are not always discovered to follow the exponential Boltzmann distribution, other than that, many sources are actually found to demonstrate power-law behaviors. In 1988, Tsallis proposed an entropy formula with a nonextensive parameter q to describe the prevalent power-law behaviors [8–11]. In the limit of $q \to 1$, it recovers the familiar Boltzmann statistics. This novel statistics is named the Tsallis statistics, or nonextensive statistics. It has several well-defined mathematical rigidities and performs impressively well in various practical domains. Empirically, the Tsallis statistics is widely employed to explain data from nonextensive viewpoints in various aspects in physics, chemistry, economics, computer science, biology, cellular automata, self-organized criticality, scale-free networks, linguistics, and other sciences (see Chap. 7 in Ref. [11] and references therein for extensively concrete examples). Theoretically, the nonextensiveness

can be explained in terms of fluctuations of temperature [12], or superstatistics [13]. The physics behind nonextensiveness is believed to be long-range correlations, strongly quantum entanglements, and deformed phase space due to insufficient statistics in space and/or time [9–11].

For the comprehensive existence of the Tsallis statistics, it appears intriguing to look into its digit distribution, besides the studied canonical ones [5]. The influence from the nonextensive parameter q on results may reveal further regularities of natural statistics. We study the first-digit distribution of nonextensive statistics in detail both analytically and numerically. We find that, analytically, in the range of $1 \leq q < 2$, it slightly fluctuates around the first-digit law in a periodical manner with respect to the logarithm of the temperature. The deviation from Benford's law is diminutive. As q varies monotonously from 1 to 2, the amplitude of fluctuation becomes smaller and smaller, and in the limit of $q \to 2$, nonextensive statistics conforms to Benford's law exactly. Hence, it explains the underlying reason why many sources from systems with nonextensiveness respect the first-digit law in an almost precise way.

The paper is organized as follows. In Sec. II, nonextensive statistics is briefly reviewed and the normalized probability density is presented. In Sec. III, we make the analytical and numerical comparisons between the first-digit distribution of nonextensive statistics and the significant digit law. Then, several relevant insights are discussed in Sec. IV, including scale invariance, base invariance, and mantissa distribution. Section V summarizes the results of the paper.

II. NONEXTENSIVE STATISTICS

Inspired by the probabilistic description of multifractal geometries, Tsallis postulated a possible generalization of entropy [8],

$$S_q = k_B \frac{1 - \sum_{i=1}^{W} p_i^q}{q-1}, \quad q \in \mathcal{R}, \tag{2}$$

where q characterizes the nonextensiveness of the considered system and $\{p_i\}$ are the probabilities associated with $W \in \mathcal{N}$ microscopic configurations, satisfying $\sum_{i=1}^{W} p_i = 1$. In the limit of $q \to 1$, S_q elegantly recovers the conventional Boltzmann entropy, $S_{q\to 1} = -k_B \sum_{i=1}^{W} p_i \ln p_i$.

The entropy in Eq. (2) can be rewritten with the help of q algebra [10, 11],

$$S_q = k_B \sum_{i=1}^{W} p_i \ln_q(1/p_i) = -k_B \sum_{i=1}^{W} p_i^q \ln_q p_i = -k_B \sum_{i=1}^{W} p_i \ln_{2-q} p_i, \tag{3}$$

where the q-logarithmic function is defined as

$$\ln_q x \equiv \frac{x^{1-q} - 1}{1-q}, \quad q \in \mathcal{R}, \tag{4}$$

and when q approaches 1, $\ln_{q\to 1} x = \ln x$.

In contrast to the conventional entropy, S_q is non-additive for two independent subsystems A and B when $q \neq 1$,

$$\frac{S_q(A+B)}{k_B} = \frac{S_q(A)}{k_B} + \frac{S_q(B)}{k_B} + (1-q)\frac{S_q(A)}{k_B} \cdot \frac{S_q(B)}{k_B}. \tag{5}$$

It is clearly seen that the deviation of q from 1 reflects the nonextensiveness of relevant systems.

Through the entropic maximizing procedure, the distribution function $f_q(E;\beta)$ can be obtained [8–11],

$$f_q(E;\beta) \propto e_q^{-\beta E} = [1-(1-q)\beta E]^{\frac{1}{1-q}}, \tag{6}$$

where the q-exponential function is defined as

$$e_q^x \equiv [1+(1-q)x]^{\frac{1}{1-q}}, \quad 1+(1-q)x > 0. \tag{7}$$

Note that when q approaches 1, $f_q(E;\beta)$ returns to the standard Boltzmann distribution proportional to $e_{q\to 1}^{-\beta E} = e^{-\beta E}$.

After normalization to unit, $f_q(E;\beta)$ is written as

$$f_q(E;\beta) = \beta(2-q)\cdot[1-(1-q)\beta E]^{\frac{1}{1-q}}, \quad 1 \leq q < 2. \tag{8}$$

For ranges other than $1 \leq q < 2$, we are not considering in this paper for reasons listed below.

(i) When $q < 1$, there exists an upper limit for the energy, $E_{\mathrm{upper}} = [(1-q)\beta]^{-1} = k_B T/(1-q)$, whose physical meaning is not well understood yet.

(ii) While $q \geq 2$, $f_q(E;\beta)$ cannot be normalized, because the power $1/(1-q)$ becomes larger than -1.

III. THE FIRST-DIGIT DISTRIBUTION OF TSALLIS STATISTICS

With the knowledge of nonextensive statistics well-prepared from the above section, we are going to explore its first-digit distribution $P_q(d;\beta)$, which now depends on the nonextensive parameter q, besides the temperature T. Utilizing the language of probability density, the likelihood to have the first digit d equals [5]

$$P_q(d;\beta) = \sum_{n=-\infty}^{\infty} \int_{d\cdot 10^n}^{(d+1)\cdot 10^n} f_q(E;\beta)\mathrm{d}E. \tag{9}$$

Combining Eqs. (8) and (9), we can get $P_q(d;\beta)$ straightforward,

$$P_q(d;\beta) = \sum_{n=-\infty}^{\infty} \left\{ [1 - 10^n \beta d(1-q)]^{\frac{q-2}{q-1}} - [1 - 10^n \beta (d+1)(1-q)]^{\frac{q-2}{q-1}} \right\}. \tag{10}$$

A nice property of the Boltzmann-Gibbs statistics and the Fermi-Dirac statistics, is still preserved [5],

$$P_q(d;10\beta) = P_q(d;\beta), \tag{11}$$

which origins from the multiplication appearance of E and β in Eq. (8). As to be explained in Sec. IV, Eq. (11) is closely related to the scale invariance of Benford's law. Therefore, we can define a new function [5],

$$P_q^\star(d;\alpha) = P_q(d;\beta = 10^\alpha), \tag{12}$$

which appears to be a one-periodical function, *i.e.*, $P_q^\star(d;\alpha) = P_q^\star(d;\alpha+1)$.

By analogy with the Boltzmann distribution [5], we expand $P_q^\star(d;\alpha)$ into Fourier series,

$$P_q^\star(d;\alpha) = \sum_{n=-\infty}^{\infty} c_n e^{i\cdot 2n\pi\alpha}. \tag{13}$$

We denote the constant component of the Fourier series of $P_q^\star(d;\alpha)$, c_0, as a new function $P_q(d)$, and through nontrivial calculations resembling Ref. [5], we can get

$$\begin{aligned} P_q(d) \equiv c_0 &= \int_0^1 P_q^\star(d;\alpha)\mathrm{d}\alpha \qquad (14)\\ &= \int_0^1 \sum_{n=-\infty}^{\infty} \left\{ \left[1 - 10^{n+\alpha} d(1-q)\right]^{\frac{q-2}{q-1}} - \left[1 - 10^{n+\alpha}(d+1)(1-q)\right]^{\frac{q-2}{q-1}} \right\} \mathrm{d}\alpha \\ &= \int_{-\infty}^{+\infty} \left\{ [1 - 10^\alpha d(1-q)]^{\frac{q-2}{q-1}} - [1 - 10^\alpha (d+1)(1-q)]^{\frac{q-2}{q-1}} \right\} \mathrm{d}\alpha \\ &= \int_0^\infty \left\{ [1 - ud(1-q)]^{\frac{q-2}{q-1}} - [1 - u(d+1)(1-q)]^{\frac{q-2}{q-1}} \right\} \frac{1}{u\cdot \ln 10}\mathrm{d}u, \end{aligned}$$

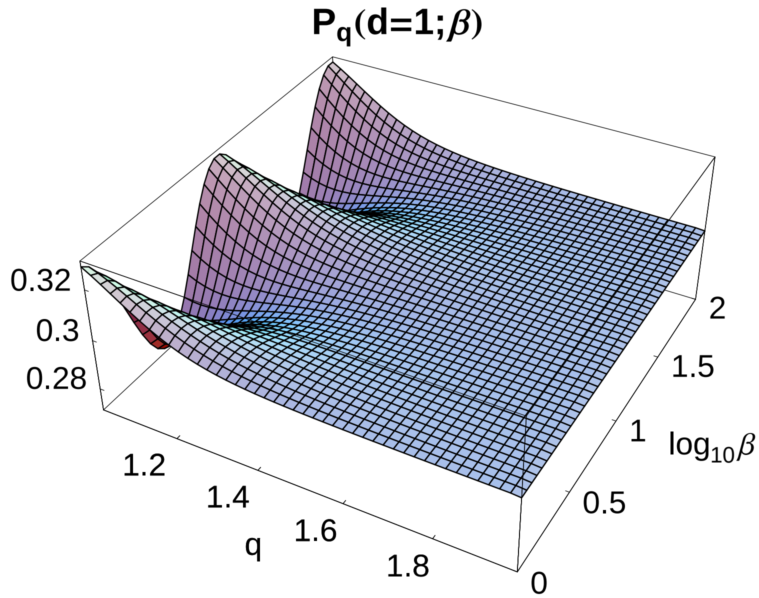

FIG. 1: (Color online). The probability of the Tsallis statistics with the first digit $d = 1$, *i.e.*, $P_q(d = 1; \beta)$, versus the logarithm of the reciprocal of the temperature $\beta = 1/k_B T$ and the nonextensive parameter q.

where a substitution $u = 10^\alpha$ is adopted. Now, the differential equation of $P_q(d)$ with respect to d is rather concise,

$$P'_q(d) = \frac{1}{\ln 10}\left(\frac{1}{d+1} - \frac{1}{d}\right). \tag{15}$$

By utilizing the normalization condition, $\sum_{d=1}^{9} P_q(d) = 1$, we attain

$$P_q(d) = \log_{10}\left(1 + \frac{1}{d}\right) \equiv P_{\text{Ben}}(d), \tag{16}$$

which turns out to be the first-digit law exactly, in analogy with three canonical statistics [5]. Therefore, we conclude that, at a fixed q, the first-digit distribution of nonextensive statistics fluctuates around the significant digit law periodically, with respect to α, or equivalently, the logarithm of β. Furthermore, we stress that, the central value, *i.e.*, c_0, is independent of the nonextensive parameter q. In contrast, as we will see later, the strength of fluctuation depends on q.

Numerically, the probability of owning the first digit $d = 1$, *i.e.*, the function $P_q(d = 1; \beta)$, is illustrated in Fig. 1. The cross section of the figure at $q = 1$ is the very case for the Boltzmann statistics [5]. The maximum amplitude to deviate from Benford's law is less than 0.03 when $q = 1$ [5]. For $q \neq 1$, from the figure, we see that the deviation is even

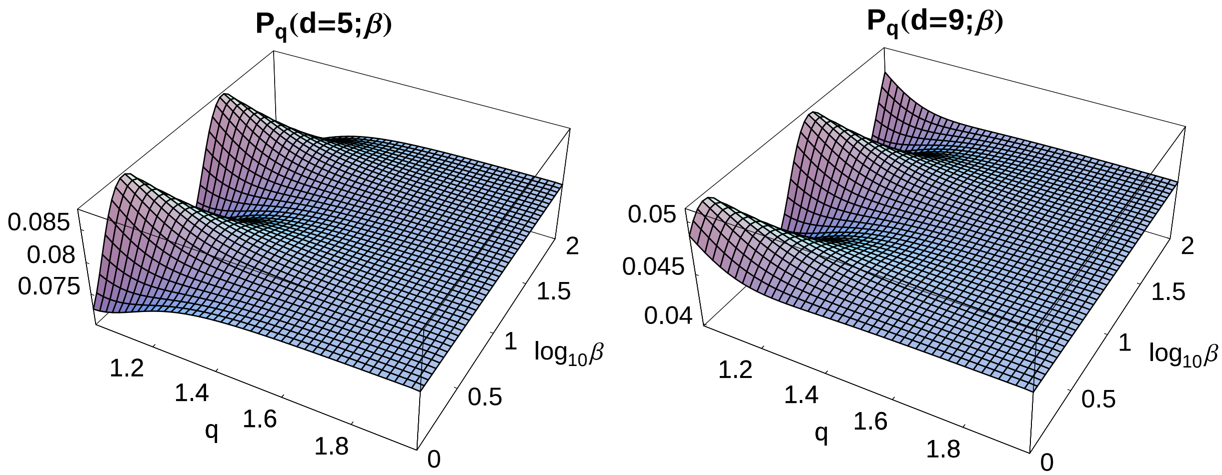

FIG. 2: (Color online). The probabilities of the Tsallis statistics with first digit $d = 5$ (left) and $d = 9$ (right), versus the logarithm of the reciprocal of the temperature $\beta = 1/k_B T$ and the nonextensive parameter q.

smaller and decreases monotonously with increasing q. In the limit of $q \to 2$, we prove in Sec. IV analytically that the digit distribution of nonextensive statistics complies with Benford's law exactly.

The probabilities for digits other than digit 1 behave similarly. In Fig. 2, we depict the probabilities of possessing the first digit as $d = 5$ and $d = 9$ as examples, versus the nonextensive parameter q and the logarithm of β. The central values follow exactly Benford's law. The deviations from Benford's law for digits 2-9 are smaller than that for digit 1. Here for 5 and 9, their deviations are both less than 0.01 for $q = 1$ [5], and further smaller for $q \neq 1$.

IV. DISCUSSIONS

Benford's law applies to numerous data from natural sources, *e.g.*, areas of lakes, lengths of rivers [7], physical constants [14], various quantities of pulsars [15], hadron full widths [16], complex atomic spectrum [17], and α-decay half lives [18, 19]. Meanwhile, nonextensive statistics performs successfully at describing ubiquitous power-law behaviors [11], and possesses elegantly conceptual merits as well [8–11]. Hence, the regularities between Benford's law and nonextensive statistics, especially, the influence from the nonextensive parameter q,

appear as an amusing and significative issue.

Generally, q is explained as a parameter reflecting microscopic mechanism of deforming phase space or fluctuating the temperature of the system, however unfortunately, due to the imperfect understanding of the dynamical details of generating nonextensiveness, it cannot be determined as a prior nowadays. A practical way to obtain q is through numerical fittings to Eq. (8). Our results might point to another possible way to access q by counting the occurrence of the first digits and investigating their deviations from Benford's law, which costs less time than conventional fitting procedure.

Now in the following, let us discuss three properties of the first digit law related to the studies presented in this paper.

First, Benford's law is scale invariant, which is discovered by Pinkham in 1961 [20]. This property means that the law remains unchanged under numerical rescalings, hence does not depend on any particular choice of units. As mentioned, in our study, it is closely related to Eq. (11), which is ascribed to the multiplication appearance of β and E. The rescaling of β is equivalent to inversely rescaling E. Therefore, concerning the scale invariance of Benford's law, we expect that Eq. (11) emerges naturally. Worthy to note that, Benford's law is also base invariant, which means that the regularity of the law is independent of the base b [21, 22]. In the binary system ($b = 2$), octal system ($b = 8$), or other base system, the data, as well as in the decimal system ($b = 10$), all fit the general Benford's law,

$$P_{\mathrm{Ben}}(d) = \log_b \left(1 + \frac{1}{d}\right), \quad d = 1, 2, ..., b - 1. \tag{17}$$

Second, the independence of $P_q(d)$ on q in Eq. (16) breaks out as an astonishing result at first sight. However, it was proven by Smith by utilizing the language of digital signal processing in his textbook that, in reference to the logarithmic coordinates, the digit distribution of any probability distribution function should fluctuate around Benford's law with respect to the rescaling parameter [23]. Therefore, the q-independence is expected. As for our study, as stated above, the dependence on β is equivalent to the rescaling of E. Therefore, the oscillation of the digit distribution around Benford's law is consistent with Smith's analysis from the viewpoint of rescaling.

The last point we would like to present is the proof of the coincidence of the first-digit distribution of nonextensive statistics and the significant digit law in the $q \to 2$ limit. It can be ascribed to the "$1/m$" behavior of mantissa distribution [5]. The mantissa $m \in [\,0.1,\ 1\,)$

is the significant part of a floating-point positive number x, defined uniquely as $x = m \times 10^n$, where n is an integer. It was pointed out that, if the probability density of m is distributed according to $1/m$, then Benford's law is guaranteed [5]. Actually, the "$1/m$" distribution of mantissa is equivalent to the n-digit Benford's law [5]. For nonextensive statistics, when q approaches 2, $f_{q\to 2}(E;\beta) \propto 1/(1+\beta E)$. The integral divergence occurs when $E \to +\infty$, where the behavior of $f_q(E;\beta)$ is proportional to $1/E$. Thus, all mantissa contribution comes from the $E \to +\infty$ region with probability density proportional to $1/m$. Consequently, nonextensive statistics follows Benford's law exactly in the limit of $q \to 2$ accordingly.

V. SUMMARY

The Maxwell-Boltzmann-Gibbs statistics is one of the most celebrated achievements in the history of physics, and it has many implications and applications in various physical as well as social domains. However, there also exist numerous examples characterized by power-law behaviors other than the canonical exponential distribution. The power-law behavior is elegantly and efficiently described by nonextensive statistics with a nonextensive parameter q, proposed by Tsallis in 1988.

While the Boltzmann statistics was proven to follow Benford's law [5], which states the uneven occurrence of the first nonzero digit, then it becomes intriguing to look into the digit distribution of nonextensive statistics and the dependence on the nonextensive parameter q. We find analytically that the first digit distribution of nonextensive statistics has similar behaviors as that of the Boltzmann distribution. It fluctuates slightly and periodically around Benford's law with respect to the logarithm of the temperature. With increasing nonextensiveness in the range of $1 \leq q < 2$, the fluctuation decreases monotonously. In the limit of $q \to 2$, nonextensive statistics follows Benford's law exactly. Furthermore, the fluctuations of temperature in physical systems can smooth down the oscillation to the central value, which corresponds exactly to the first-digit law. Therefore, we reveal that, the frequent appearance of Benford's law in the natural and social data is theoretically expected in systems that follow nonextensive statistics.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (Nos. 11005018, 10721063, 10975003, 11035003). It is also supported by Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (Chun-Tsung Endowment) at Peking University, and by National Fund for Fostering Talents of Basic Science (Nos. J0630311, J0730316).

---

[1] J.C. Maxwell, Phil. Magazine **19**, 19 (1860).

[2] J.C. Maxwell, Phil. Magazine **20**, 21 (1860).

[3] L. Boltzmann, Wien. Ber. **76**, 373 (1877).

[4] J.W. Gibbs, *Elementary Principles in Statistical Mechanics*, Yale University Press (1902).

[5] L. Shao, B.-Q. Ma, Physica A **389**, 3109 (2010), arXiv:1005.0660 [physics.data-an].

[6] S. Newcomb, Am. J. Math. **4**, 39 (1881).

[7] F. Benford, Proc. Am. Phil. Soc. **78**, 551 (1938).

[8] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[9] C. Tsallis, Braz. J. Phys. **29**, 1 (1999), arXiv:cond-mat/9903356.

[10] C. Tsallis, Eur. Phys. J. A **40**, 257 (2009), arXiv:0812.4370 [physics.data-an].

[11] C. Tsallis, *Introduction to Nonextensive Statistics Mechanics – Approaching a Complex World*, Springer, New York (2009).

[12] G. Wilk, Z. Wlodarczyk, Phys. Rev. Lett. **84**, 2770 (2000), arXiv:hep-ph/9908459.

[13] C. Beck, E.G.D. Cohen, Physica A **322**, 267 (2003), arXiv:cond-mat/0205097.

[14] J. Burke, E. Kincanon, Am. J. Phys. **59**, 952 (1991).

[15] L. Shao, B.-Q. Ma, Astropart. Phys. **33**, 255 (2010), arXiv:1005.1702 [astro-ph.IM].

[16] L. Shao, B.-Q. Ma, Mod. Phys. Lett. A **24**, 3275 (2009), arXiv:1004.3077 [hep-ph].

[17] J.-C. Pain, Phys. Rev. E **77**, 012102 (2008), arXiv:0801.0946 [quant-ph].

[18] B. Buck, A.C. Merchant, S.M. Perez, Eur. J. Phys. **14**, 59 (1993).

[19] D. Ni, Z. Ren, Eur. Phys. J. A **38**, 251 (2008).

[20] R.S. Pinkham, Ann. Math. Stat. **32**, 1223 (1961).

[21] T.P. Hill, P. Am. Math. Soc. **123**, 887 (1995).

[22] T.P. Hill, Stat. Sci. **10**, 354 (1995).

[23] S.W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, available at `http://www.dspguide.com`.